



Consolidation or multiplicity in supply logistics for health commodities?

Ebenezer Kwabena Tetteh

Department of Pharmacy Practice & Clinical Pharmacy, School of Pharmacy, College of Health Sciences, University of Ghana, Legon, Accra, Ghana



ARTICLE INFO

Article history:

Received 27 September 2021

Received in revised form 1 December 2021

Accepted 11 January 2022

Keywords:

Consolidation

Distribution

Logistics

Multiplicity

Supply

Wholesaling

ABSTRACT

Public health-sectors of most low- and middle-income countries (LMICs) run a central medical stores (CMS) model that is monopolistic in character. Concerns raised about monopolistic CMS arrangements stress the need to encourage cost-reducing efforts and improve service levels (outputs) by having multiple competing logistics institutions. This paper examines the desirability of consolidation or multiplicity in supply logistics by focusing on the task of inventory management (that is, distribution, storage and warehousing). The paper uses theory and historical evidence to describe and suggest a desired form of multiplicity in LMICs. Consolidation shouldn't progress to the point of monopoly and multiplicity doesn't mean having an infinite number of logistics institutions. A limited number (2–10) of logistics institutions, that are full-line and national in their scope and scale of operations, should be enough to provide choice, support competition and minimize the risk of supply disruptions. Health policy and planning in LMICs should explore ways of turning existing logistics institutions in the public, private and non-governmental sectors into a multiplicity of types that are capable of assuring uninterrupted supplies of health commodities.

1. Introduction

Public sectors of most low- and middle-income countries (LMICs) rely on a central medical stores (CMS) supply chain model in their efforts to assure health commodity security. This CMS model often describes a monopolistic logistics system with no alternative competitors to encourage cost-reducing efforts and improved service or output levels. Yet the apparent dominance of CMS monopoly in the public-sector, and conventional wisdom advocating integration and/or consolidation of logistics systems doesn't fit well with reality. What is commonly observed is a hybrid state with multiplicity of government, non-governmental and private logistics systems, not integration or consolidation.^{1,2} The benefits of multiplicity, it is argued, are: (1) greater flexibility to maintain supply with alternative options, (2) competition to lower costs and improve service levels, and (3) opportunities to tailor logistics systems to meet priority health program objectives and specific product needs. Multiplicity, however, has the following downsides: (1) the need for increased oversight [regulation] and coordination, (2) perverse incentives to collude and increase costs, and (3) overspecialization on select geographical regions, health facilities or commodities can reduce supply flexibility.³ Clearly, how logistics systems are designed and organized in LMICs will determine whether the benefits of multiplicity (consolidation) outweighs the downsides. Beyond this general statement, the benefits and downsides listed provide no indication of what the desired form of multiplicity is or should be. That aside, there are several definitions of multiplicity.

Health systems in LMICs can generally be described as having three sectors: government, non-governmental and private. Multiplicity could mean having one or more (competing) logistics systems serving each sector separately or serving all three sectors together. We could also have a single

monopolistic logistic system serving each sector but this will be counted as multiplicity. Besides the case of a single logistics system serving all three sectors, a blend of arrangements is possible. We could have monopolies in government and non-governmental sectors but two or more logistics systems in the private sector. Such an arrangement will be counted as multiplicity. Within each sector, one may observe multiplicity simply because existing and emerging logistics systems only serve health facilities within some geographical locations. Or, they supply only a subset of health commodities, for e.g., family planning products or just essential medicines. These examples are, however, not the only forms of multiplicity. A functional logistics system is made up of institutions undertaking the following logistics tasks: product selection, (demand) quantification and procurement, inventory management (warehousing, storage and distribution) and service delivery. Since what is needed for commodity security is synchronized execution of these logistics tasks, multiplicity could mean having multiple logistics institutions undertaking one or more logistics tasks on behalf of affiliated actors in all three sectors. It is also important to make a distinction between (1) multiplicity in relation to the number of logistics institutions (units), where each unit is a subsystem of the logistics system; and (2) multiplicity *within* a logistics institution, i.e., multiple subunits or divisions that form part of a single logistics institution (unit). This is often observed for logistics institutions undertaking the task of inventory management. A network of subunits could, for e.g., be organized into echelons or tiers (most commonly central, regional or district levels) to reflect public governance and administrative boundaries within a country. If logistics systems can be designed in various ways that fit with the label multiplicity, what then is the desired form of multiplicity for supply logistics?

According to Bornbusch and Bates,³ the challenge for LMICs is identifying a state of “prudent multiplicity” where the incremental costs of having

one more logistics institution or system is worth the incremental benefits. The benefits being promoting competition and choice; hedging supply-disruption risks to assure commodity security, and better health outcomes. This paper addresses that challenge by focussing on institutions executing the logistics task of inventory management (= storage, warehousing and distribution). This is because most statements about “supply logistics” are in fact references to the task of inventory management. Also, because a logistics system is the sum of its component parts and multiplicity can take different forms, it is more informative to evaluate each logistics task separately from the others. For instance, the national electronic procurement service in Chile (ChileCompra or “ChileBuys”), discussed by Bornbusch and Bates,³ is an example of multiplicity to improve the logistics task of procurement. This e-procurement service was adopted in response to decentralization and the autonomy regional health authorities had to bypass pooled procurement run by Chile's CMS. With an electronic procurement platform, regional health authorities in Chile can still bypass the CMS but they have to procure directly from the same set of suppliers the CMS deals with. The Chilean case shows how multiplicity can help maintain, in a decentralized environment, lower prices and cost savings achieved through pooled procurement. What it doesn't show is the benefits of multiplicity in inventory management. In other words, a valid overall assessment of the desirability of multiplicity of logistics systems must first evaluate the desirability of having multiple institutions executing each logistics task. Health planners in LMICs can then use these assessments to build functional logistics systems with the form of multiplicity most suited to their country contexts.

In what follows, this paper examines whether multiplicity in supply logistics (inventory management) is desirable. The paper has two aims. One is to identify and characterise a form of multiplicity that is best suited for assuring uninterrupted supplies of all health commodities needed. Two is to show that the downsides of multiplicity (consolidation) are not universal. Much depends on whether logistics systems in LMICs can be organized, reconfigured or transformed in ways that balance the benefits of consolidation and multiplicity.

2. What is the desired form of multiplicity?

To determine whether multiplicity in supply logistics (inventory management) is desirable, and in what form, this paper makes the following assumptions. First, each LMIC market is made up of three sectors or submarkets: public, private sector and non-governmental. In each of these submarkets, health facilities act as price-sensitive buyers or there is a strategic purchaser acting on their behalf. Second, health planners have identified and selected a basket of essential health commodities that have positive clinical and economic value in treating the range of diseases that the general population might suffer from. Health planners' objective is to assure aggregate commodity security – which is defined as the uninterrupted supply of all health commodities in the essential basket whenever and wherever these commodities are needed. Third, health planners desire to have logistics institutions and for that matter logistics systems that are designed for “last mile distribution” of the basket of selected health commodities in whatever quantities deemed appropriate, constrained only by the number and geographical spread of health facilities. Fourth, health planners have no inherent dislike for relying on private logistics institutions (private wholesalers) to undertake the logistics task of inventory management. Fifth, given the focus on inventory management (what is called pure wholesaling in private markets), the output dimensions health planners care about are those shown in Table 1 below. The dimensions in Table 1 are naturally metrics to measure how well inventory management is executed. Given these assumptions, the paper describes in detail characteristics of what it considers as the desired form of multiplicity. These characteristics are: (1) logistics institutions with long-lives, (2) competition and choice; and (3) hedged supply risks.

Table 1

Dimensions of output.

Logistics function (task)	Dimensions or indicators of output
Inventory management: Warehousing, storage and distribution	Percentage of commodities with adequate shelf-life Percentage of logistics facilities holding stock within minimum and maximum levels Percentage of inventory counts that matches records Frequency of stockouts and/or emergency orders Average duration of stockouts Percentage of logistics facilities in compliance with guidelines Percentage of stock expired or damaged (accident rates) Percentage of orders that are filled as requested (order fill rate) Distribution lead times (versus the average) Percentage of complete reports Percentage of complete reports submitted on time Ratio of transportation cost to value of commodity Total warehousing, storage and/or distribution costs Percentage of markup on commodities (in a cost-recovery system)

Source: USAID.^{27,28}

2.1. Logistics institutions with long lives

To ensure, today and tomorrow, uninterrupted supply of health commodities in all therapeutic categories, whenever and wherever they are needed, health planners must be concerned about the long-run survival of logistics institutions, whether public, private or non-governmental. This point is best explained using mathematical expressions.

Let t refer to each short-run time period, d is the number of districts or regions within the country, n is the number of health facilities within each district or region, c is the number of therapeutic categories and j is the number of essential health commodities in each therapeutic category. To ensure long-run survival of a logistic institution undertaking the task of inventory management, the following conditions shown in Eqs. (1), (2) must be met:

$$\left\{ \sum_{t=1}^T \sum_{d=1}^D \sum_{n=1}^N \sum_{c=1}^C \sum_{j=1}^J (P_{jndt}^i - C_{jndt}^i) * Q_{jndt}^i \right\} \geq F + \text{Profit} \quad (1)$$

$$P_{\text{enduser}}^i = P_{\text{acquisition}}^i + P_{jndt}^i \quad (2)$$

where i refers to a given logistics institutions (or group of institutions), P is the price charged for a commodity or service provided, C is the corresponding marginal or incremental costs (approximated by average variable costs), and Q is quantities of health commodities supplied (a product of number of filled requisition orders and the number of SKUs per filled order). F refers to the sequence of repeated sunk or (quasi)fixed cost investments made in setting up, maintaining or upgrading the needed logistics infrastructure (plant). Such sunk or fixed costs (term F) will cover building warehouses or distribution centres with each having adequate power supply, lighting, spacing [floors and docking areas for a fleet of delivery trucks and vans etc.]; security, fire and safety devices and measures [first-aid, tracking and antitheft kits etc.]; storage facilities for commodities that are inflammable or require strict temperature control. F will also include the costs of furnishing warehouses with equipment for handling commodities (pallets, pallet racks, static shelves, lift trucks, rolling warehousing ladders etc.) as well as manual or computerized logistics management information systems (LMIS) linked to devices for automatic data collection (bar code

readers, radiofrequency identification technologies etc.). It will also include continuing or repeated sunk or fixed costs made (in LMIS, warehouse capacity and transportation) to improve outputs (lower distribution lead times etc.). It will be necessary to do so with availability of a broader range of health commodities and increasing demands (for e.g. increases in number and geographical spread of health facilities). Without these continuing or repeated investments, commodity stockouts will be more frequent.

Eq. (2) suggests the final price paid by each end-user, customer or service delivery point, in each time period must be the sum of commodity acquisition prices and prices for inventory management. Prices for inventory management, in the short-run, are what is commonly referred to as distribution markups. These prices must be estimated taking into account prevailing LMIC demands and not just the continuing or repeated costs required to assure uninterrupted supplies. Each functional logistics institution must then aim to secure over time a sequence of short-run prices that allows recovery or recoupment of all fixed or sunk costs plus a competitive normal profit, where relevant. If health planners prefer a policy of zero distribution markups (zero prices for inventory management), then some other source of revenue is called for to cover costs.

2.2. Competition and choice

If logistics institutions are to have long-lives, they should be able to cover all the economic costs involved in assuring commodity security. This however should not come at the cost of monopolistic inefficiencies. In the absence of competitors, the only constraint on a monopoly in executing the logistics function of inventory management is zero demands which is at odds with the objective of assuring commodity security. Competition and choice is desirable for a number of reasons. First, it reduces how much of societal resources is devoted to inventory management. In contrast to the indeterminacy of pure bargaining, competition helps reveal a wider range of feasible supply prices as suppliers (in our case logistics institutions) strive for incremental business volumes or market shares. Compared to a reliance on hard (take-it-or-leave-it) bargaining skills, competition driven by price-elastic demands is akin to the bargaining tact of taking business elsewhere. Faced with price-sensitive purchasers, logistics institutions will give serious consideration to cost-reducing process innovations.⁴ A number of preconditions must however be met or created to derive the societal benefits of competition.

One, there must be an adequate (not an infinite) number of competing units offering a homogenous or near-homogenous service levels (outputs). That is, the competing units must be similar, and carry out the same set of logistics activities. Since the planner's objective is to assure uninterrupted last-mile deliveries of all health commodities, whenever and wherever they are needed, the competing units (logistics institutions) should be designed for and capable of fulfilling that task. Two, unlike theoretical ideals of perfect competition where prices charged must be equal costs, short-run prices for inventory management (the pure wholesaling function) can be higher than the corresponding marginal or incremental costs. This is in fact optimal and necessary in the presence of (continuing or repeated) fixed and sunk costs, although the optimal deviation from incremental or marginal costs depends on how price sensitive purchasers are.⁵ A simplified but useful representation of real-world imperfect competition is the generalized "many-firms" Cournot model.^{6,7} In this model, the excess of prices over costs is given by Eqs. (3), (4).

$$L^i = (P_{jndt}^i - C_{jndt}^i) / P_{jndt}^i = S^i / \eta \quad (3)$$

$$L = \sum_{i=1} (S^i)^2 / \eta = HHI / \eta \quad (4)$$

where L^i is the Lerner index (a measure of deviation of prices from costs) for a given firm (logistics institution), S is the share of the aggregate demand facing a given logistics systems or supply chain, L is the sum of quantity- or share-weighted Lerner indices, HHI is the Hirschman-Herfindahl index

and η is the price-sensitivity of demands of health facilities or the strategic purchaser acting on behalf of health facilities (in each sector).

Eqs. (3), (4) suggest (1) policies, guidelines, regulations or rules that encourage collusion between the competing logistics units and (2) a very small number of logistics institutions will generate a sequence of higher prices over the relevant time periods than is socially optimal or necessary. Conversely, a higher number of logistics units will not lead to competitive pricing without purchaser price-sensitivity. Thus price-sensitivity by healthcare purchasers' demands is far more important in providing adequate market discipline and incentives for efficiency. To be specific, what is needed is at least two competing logistics institutions with near-homogenous outputs (see Table 1) plus price-sensitive purchasers. Being price-sensitive means being able to credibly exercise the bargaining threat of taking business elsewhere such that lower prices for inventory management are rewarded with more than equiproportionate increase in demand volumes or market shares.⁸ Striving for custom (incremental demand volumes or market shares) is the hallmark of real-world imperfect competition⁴ – and it can easily be translated into a standard operating procedure. It certainly doesn't mean giving all business or contracts to the logistics unit offering the lowest price; just that prices should be inversely related to demand or business volumes. See van Valen et al.⁹ for a matrix that can be used to share demand or business volumes among multiple winners (competitors). The expectation here is: competition will get logistics institutions to make the continuing or repeated sunk-cost investments needed to maintain excellent outputs and provide adequate incentives to lower the costs involved either by adopting cost-reducing process innovations, wider short-run spreading of fixed sunk costs over larger demand volumes and exploitation of long-run economics scale and scope. Competition spurred on by price- and output-sensitive purchasers will initiate and enforce constant experimentation and a search by logistics institutions for the optimal scale of operations, especially with regards to the number and size of distribution centres. Multiple competing logistics institutions (units) should put the brakes on unrestrained multiplicity of distribution centres (subunits) *within* each unit.

We want to emphasize that, to assure secure supplies of all health commodities in the basket of essentials, the competing units (public, non-governmental or private logistics institutions) must be full-line and national, not full-line subnational or short-line national or short-line subnational entities. Full-line here means the competing units supply, at least, everything included in the basket of essential health commodities. These units must be national (not nationalized) in that all health facilities including those in remote areas have secure lines of supply. For competition to thrive, none of the competing units should have exclusive (geographical) distribution or marketing rights for any of the health commodities included in the basket of essentials. That is, full-line national logistics institutions must be protected from competing units that engage in what is known as cream skimming and skimping. Cream skimming arises from a specialized focus on, for e.g., fast-moving, profitable health commodities and/or on serving only selected districts or regions within any LMIC, perhaps those with higher relative incomes. Skimping refers to attempts to avoid the continuing or repeated sunk-cost investments required to maintain outputs at desired levels. For instance, a competing logistics institution might opt not to supply "unprofitable" vaccines requiring stricter temperature control to avoid investing in cold-chain storage facilities. Or, instead of direct (last mile) delivery, it will only offer customer pickup arrangements.¹ The issue is: under imperfect competition, logistics units might not necessarily earn zero excess profits: there will be some winnings (winners making excess profits) and losses (losers making less than normal profits) even though

¹ There are two instances where customer pickup arrangements make sense. One, logistics institutions cannot provide or health facilities are not willing to pay for direct (last mile) delivery at *any price* even if this means higher order fill rates, lower direct costs and avoiding the indirect time costs of healthcare workers travelling to pick up orders when they should be caring for the sick. Two, close proximity to distribution centres makes it possible for health facilities to pick up their orders at costs that are lower than the prices charged for inventory management. It is hard to imagine the former and the latter is unsatisfactory since that argument does not hold for health facilities located in remote regions.

firm-specific (market-wide) profits maybe close to zero. Cream-skimming and skimping creates opportunities for logistics units will smaller scale or scope of operations to erode “profits hidden beneath the zero” that are crucially needed for long-run survival of competing full-line national units.⁵

Cream-skimming and skimping should be barred or at least discouraged since these practices work against the objectives of ensuring aggregate commodity security and having long-lived logistics institutions. The sequence of fixed or sunk-cost investments required for direct (last mile) delivery of all health commodities to all health facilities in different regions in any given LMIC could be in thousands or millions of dollars. But these costs spread over larger demand volumes and multiple time periods will be comparatively small. In a number of scenario analyses of direct (last mile) delivery of vaccines in Guinea, Madagascar and Niger, Prosser et al.¹⁰ report that, although the sum of additional start-up and operational costs were in millions of US dollars, this amounted to less than \$1 per dose of vaccine delivered over a 1-year period. As shown in Eqs. (1–2), each competing logistics units has to secure a sequence of prices low enough to allow full recovery of all (continuing or repeated) sunk costs made over the long-lives of these institutions. Prices charged for inventory management then need not be frontloaded in attempts to recover the costs incurred within the shortest possible time. The costs involved should be recovered over the long-run, not after a few years. Considering low household incomes and low per capita expenditures on healthcare in LMICs, health planners should consider putting limits on the number and type of competing logistics units licensed to operate. Barring overspecialization (skimping and/or cream-skimming), for e.g., means competitive entry will be observed only by full-line national entities. This together with limits on the number of logistics institutions will consolidate larger business volumes (Q) for each competing unit.¹¹ Indeed, to consolidate even larger business volumes (and a steady stream of cash flows to support capital investments), competing logistics institutions, in LMICs, should serve all three sectors: public, private and non-governmental. This is particularly important where the costs of raising capital is high. In addition, volume guarantees may be considered but only as the outcome of a competitive process in which the size of these guarantees are inversely related to prices charged for inventory management.

2.3. Hedged supply risks

In Section 2.2, we argued that competition and choice only needs at least two full-line, national logistics units. The question is: are these limits on multiplicity desirable in a real world with diverse risks and uncertainty, where a lot can go wrong or deviate from expected outcomes. Instead of a stable source of supply, one could have variable outputs (throughput or service levels), issues with quality of health commodities and unreliable suppliers. These are the opposites of what characterises stable supply (dependable lead times, stable outputs, infrequent breakdowns, reliable suppliers, diverse set of suppliers and minimal constraints on supply capacity).^{12,13} In this section, we examine whether an infinite number of (competing) logistics institutions is necessary for mitigating supply risks. We do so by assuming each competing logistics unit has the same prior probability of a breakdown but these probabilities are independent. If these probabilities are only conditional on LMIC context, then the joint probability that all competing units will breakdown is ΠProb_i , where i refers to each logistics unit in a set of N units. Conversely, the probability of assuring commodity security is $1 - (\Pi \text{Prob}_i)$. If there are two logistics units with prior probabilities of a breakdown equal to 0.1, then there is only 0.01 ($= 0.1^2$) chance that *both* logistics units will breakdown. On the other hand, if the prior probabilities are 0.9, there is a 0.81 ($= 0.9^2$) chance that both logistics units will breakdown. If the prior probabilities are 0.6, one will need ten competing logistics units to attain a 0.99 chance of assuring commodity security. This outcome can be achieved with just three units if prior probabilities are 0.2. These simplified examples suggest the benefits are multiplicity are lower when the probabilities of breakdown approaches one.

Generally, one's ability to hedge against supply-risks via multiplicity diminishes as the number of logistics units increase. But it does so more

quickly when there is more than a 50:50 chance of a supply breakdown. Fig. 1 provides a graphical illustration for a selected set of prior probabilities of supply disruptions. The picture changes somewhat if the prior probabilities of a breakdown or supply interruptions are dependent. If prior probability of a supply-breakdown is 0.1 for one logistics unit but for another the probability of a breakdown conditional on the first unit breaking down is 0.4 (perhaps because both units rely on the same set of institutions undertaking one or more logistics functions other than inventory management), the joint probability of both logistics units breaking down will be higher. That aside, the probability of a single logistics unit experiencing a disruption will, in part, be determined by the number of subunits or divisions (in each echelon or tier). The reason is: for each logistics unit to be functional, all of its component parts (subunits) must be functional. If the number of subunits is 2 and the probability of each subunit experiencing a fault is 0.1, then the probability of the logistics institution being functional (as a sum of its component parts) will be 0.81. With 12 (30) subunits, the probability that the logistics unit will remain functional is 0.28 (0.04). So besides cost inefficiencies, multiplicity *within* a logistics unit makes it more difficult to achieve a 100% order fill rate, for example. The additional benefits of risk-hedging via multiplicity doesn't just diminish with increases in the number of logistics institutions, but also with interdependent risks and increases in the number of subunits (per tier or echelon) within each logistics institution.

Whilst these observations hold in most cases, the answer to the question of whether multiplicity is desired also depends on health planners' aversion to supply risks and their prior subjective estimates of the probability of disruptions; and whether these expectations are confirmed or refuted using historical data or events experienced. Based on Fig. 1, we believe 2–10 (competing) logistics units should be adequate for risk mitigation. A more risk-averse health planner, however, may prefer more than 10 logistics units even though the incremental reduction in disruption risk with each additional unit gets smaller and smaller. Note also the magnitude (of the consequences) of a supply breakdown is as important as the risk or probability that it occurs. Rare (low probability) catastrophic events that lead to a steady fall in order fill rates to zero or complete exhaustion of stock on hand (a fire outbreak for e.g.) can be adequately hedged via multiplicity. Effective internal, day-to-day management of logistics activities should be enough to hedge against low and high probability events with minor impacts (such as small variations in demand, staff absenteeism, work-related injuries, staff strike actions). High probability high impact (catastrophic) events, on other hand, cannot be adequately hedged via multiplicity. In such situations, it is more difficult to assure commodity security since what is needed is frequent and costly redesign or reconfiguration of logistics units.¹⁴

3. Parallels with historical evidence

From the above, the desired form of multiplicity in supply logistics in LMICs should be one characterised by a limited number of competing full-line, national (not necessarily nationalized) logistics institutions doing business with price- and output-sensitive purchasers in all three sectors. This section draws parallels between that proposition and how private pharmaceutical wholesalers in Europe and the United States (US) have evolved over time.

3.1. Pharmaceutical wholesaling in the US

Over the period 1700–1929, there were in the US mostly small short-line regional wholesalers located near seaports since this made the business of importation easier. Settling in new territories and new transportation routes (roads and canals) encouraged entry by broker-middlemen wholesalers who had to travel from the interior of the country to port cities in order to procure their products and/or assess the quality of products purchased. Growth in numbers of wholesalers including broker-middlemen was fuelled by growth in healthcare demands. By 1929 there were approximately 55,000 drug stores compared to 25,000 drug stores in 1880,

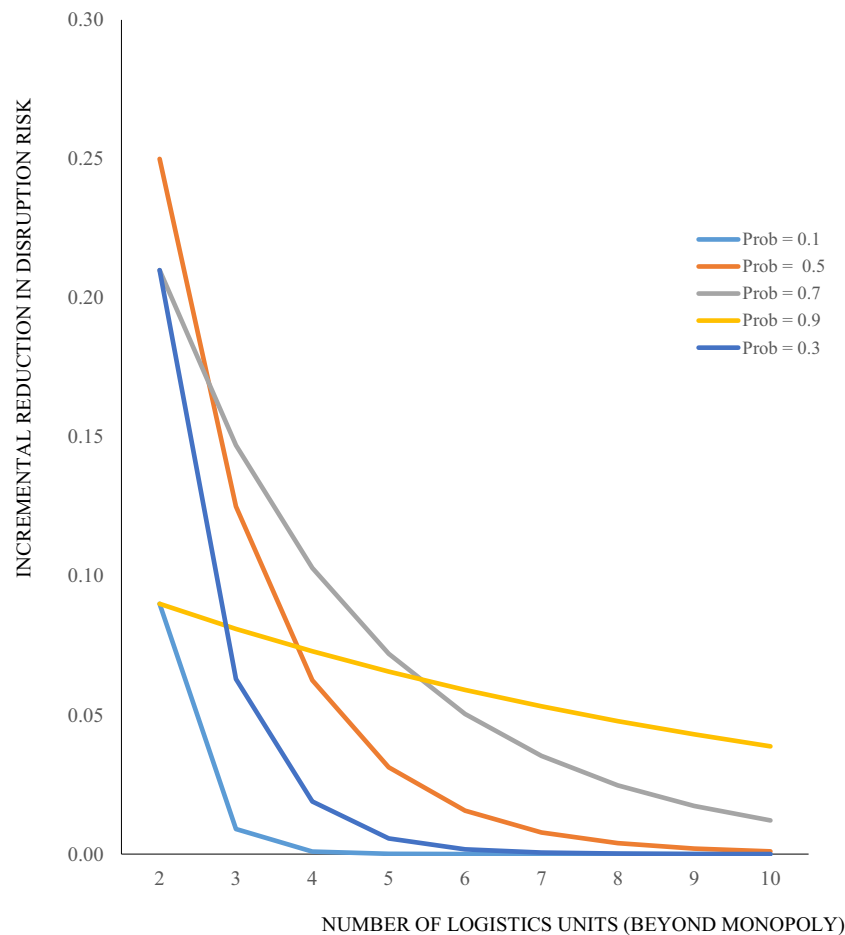


Fig. 1. Limits to hedging supply disruption risk via multiplicity. *Notes:* The incremental (marginal) reduction in risk of disruption was computed as the difference between the probabilities of assuring commodity security ($1 - (I\text{Prob}_i)$) as the number of (competing) logistics units increase by one. The starting point is the case of monopoly, and it is assumed the prior probability of any logistics unit breaking down is independent of other units.

although the number of drug stores relative to population size remained at roughly one store for every 2000 people. In addition, the number of hospitals increased from 178 in 1872 to 4000 in 1910. Competition between the broker-middlemen was however confined to geographically distinct markets (territories). Fears of excessive and unfair competition led to the creation of the National Wholesale Drug Association (NWDA) in 1882. The NWDA had 201 active members in 1886. The period 1929–1978 saw entry by the first national wholesaler, McKesson & Co., after mergers of the parent company in 1929 with 64 other wholesalers in 31 cities. Given increasing healthcare demands and fierce price competition between broker-middlemen, the emergence of a national wholesalers can be seen as an attempt to exploit in the long-run economies of scale and scope in order to survive competitive pressures. Forming a national wholesaler out of smaller regional wholesalers helped avoid some or all of fixed or sunk costs (capital investments) needed to enter new previously unserved markets. It also helped build a large customer base (as it is much easier to capture loyal customers of regional wholesalers through mergers or acquisitions than trying to win over their loyal customers as a separate business unit). By 1961, McKesson served 33,000 drug stores and 5000 hospitals. Over the same period 1929–1978, there was also the emergence of other large wholesalers (formed through acquisitions and mergers of small regional wholesalers). These larger national wholesalers including McKesson switched from short-line to full-line distribution – selling both pharmaceutical and non-pharmaceutical products.^{15,16,17;18}

Over the period 1978–1996, the number of private wholesalers fell dramatically. In 1970, there were 144 wholesalers with 372 distribution centres (145 wholesalers with 395 distribution centres in 1975) but between

1975 and 1995, the number of private wholesalers dropped from 145 to 63 with 224 distribution centres. By 1996, there were 55 wholesalers with 233 distribution centres. In the year 2000, there were fewer than 50 private wholesalers with the big three accounting for 88% of the US wholesaling market. The dramatic reduction in multiplicity was mostly due to wholesaler mergers and acquisitions for the purpose of serving previously unserved geographical markets and increasing sales at existing locations. But consolidation across geographical markets, and the incentive for wholesalers to form and operate national distribution networks in the US was also given by the emergence of price-sensitive, cost-conscious hospital or retail pharmacy chains or large group purchasers acting on behalf of individual hospitals or retail pharmacies located in the different geographical regions. These healthcare purchasers (providers), themselves were faced with incentives to be cost-conscious through the introduction of prospective payment systems and managed care competition. The various cost-containment practices adopted by healthcare purchasers made the demand curves facing wholesalers more price elastic. And because these healthcare purchasers (providers) were located in different regions, their selected prime vendors were national full-line wholesalers. Regional or local wholesalers could not supply all health facilities that were located in multiple geographical regions and belonged to the same purchasing group.^{15–18}

Not all (smaller) wholesalers responded to these incentives to create national distribution networks, but the ones that responded were able to augment economies-of-scale and scope with cost savings achieved via the introduction of cost-reducing process innovations. For example, the use of order pick labels, barcodes and lightweight wearable computers to locate stocks in large warehouse space and for sorting: “night picking” of products

in warehouses to ensure orders were ready for delivery in the morning of scheduled delivery dates; and batch-order picking, i.e., picking all products listed on a given batch of orders received together so as to reduce the repeated tasks of picking the products ordered individually (thereby reducing intra-warehouse travel time per order item). These private wholesalers also offered value-added service: electronic ordering and managing computerized logistics information systems.¹⁵

3.2. Pharmaceutical wholesaling in Europe

Traditional private wholesaling in Europe is one that relies on full-line logistics units to supply a broad range of products. According to Walter et al.,¹⁹ full-line wholesalers bundle, on average, medicinal products of roughly 18 manufacturers together per delivery, selling 74% of all medicines in Europe to mostly retail pharmacies (93%), hospital pharmacies (4%), dispensing doctors (2%) and drug stores (1%). These private full-line wholesalers deliver within 3 h, 16 times a week on average.¹⁹ However, the full-line operations of these private wholesalers is in part determined by regulations of the mix of active therapeutic molecules and products containing these molecules on a country's market. In Norway, known to strictly regulate the mix of products available, full-line wholesalers in the mid-1990s supplied 4000 products containing 800 active therapeutic molecules. In Germany, full-line wholesalers supplied 70,000 products containing more than 3000 active therapeutic molecules because less strict regulations on product mix.²⁰ As in the US, private wholesaling in Europe went through phases of consolidation or erosion of multiplicity. In France, for e.g., the number of wholesalers fell from more than 150 at the start of the 1960s to 13 regional and national full-line wholesalers in 2002. In 1989, there were 2 wholesalers in Norway, 2 in Sweden, 4 in Denmark, 7 in Ireland, 20 in France, 25 in Germany, 30 in the Netherlands, 33 in the United Kingdom, 60 in Belgium, 210 in Spain and 290 in Italy. In 2001, there were 346 full-line wholesalers in the EU compared to 600 in the early 1990s, with 70% of these wholesalers based in Italy and Spain.^{20,21}

Consolidation or erosion of multiplicity in Europe was driven by a confluence of factors: (1) creation of a single European market; (2) traditional full-line wholesalers facing increased competition from foreign wholesalers, short-line wholesalers, retail pharmacy cooperatives integrating backwards into wholesaling; and (3) reduction in gross operating margins – as healthcare purchasers, who had to work with tight budgets, pursued cost-containment strategies.^{20,21} Notably, creation of a single European market led to cross-country specification of minimum standards for pharmaceutical wholesaling – and the costs of meeting these regulatory standards were more easily borne by partnering wholesalers than by small-size short-line independent ones. Full-line wholesalers who chose to be local (national) faced declining demands in contrast to pan-European wholesalers. The reduced gross operating margins of these national full-line wholesalers led to poor quality in distribution services (mostly in term of lead times and frequency of deliveries). Private wholesaler competition in Europe was such that some traditional full-line wholesalers, stocking a broad range of health commodities suffered losses. And, because 10% of all products distributed by full-line wholesalers, accounted for 85% of the profits, this created perverse incentives to switch from full-line to short-line wholesaling.^{20,21} But cream-skimming (i.e., supplying only a select range of fast-moving profitable products) does not sit well with effective competition in supplying *all* health commodities needed. Indeed, more recent literature report that short-line wholesaling is prohibited in France and Spain.¹⁹

Note that cream skimming wasn't the only way European wholesalers responded to low profit margins. They (1) increased their productivity by improving the accuracy of demand forecasts, upgrading transport infrastructure, automating warehousing activities; (2) increased cost-efficiency by closing down distribution centres; and (3) expanded their business portfolios by diversifying into non-traditional goods and services, for e.g., supplying non-drug products [cosmetics, toiletries, video tapes etc.], supplying new healthcare providers and providing financial services.²⁰

4. The importance of similarity

In Section 2, this paper argued that, to assure aggregate commodity security, competing logistics units undertaking the task of inventory management must be full-line, national entities. The word national refers generally to the geographical or administrative boundaries of the relevant market, within which the intended service delivery points including health facilities are located. In Section 3, the paper provided historical evidence from the US and Europe showing the importance of government regulation and price-sensitive healthcare purchasers (providers) in generating high-powered incentives to lower costs. In this section, we stress the importance of ensuring competing logistics institutions are similar or homogenous in terms of their outputs or services provided (i.e., there is no room for cream-skimming and skimming). We do so by looking at challenges of assuring health commodity security in China's Hubei province. In what follows, we consider the set of primary healthcare facilities within the boundaries of Hubei province in China as the relevant ("national") market.

In Hubei province, public-sector competitive tendering involves a two-envelope system under which suppliers of health commodities send bids for the procurement of health commodities whilst private distributors (wholesalers) send bids for the business of delivery (inventory management). Suppliers then select their preferred distributors from the set of bid winners. Final price to the end-user is as shown in Eq. (2). A supplier could select *any* number of distributors out of a total of some 305 (in 2015). This is not necessarily a bad policy: it only requires prices for inventory management to be inversely related to business volumes and that selected logistics units offer near homogenous outputs. The problem was out of the 305 distributors in Hubei province (most of which are only licensed to operate within a particular county), only one or two of these distributors could serve *all* primary healthcare facilities in the province. A large number of logistics units competing in smaller submarkets (counties of the province) meant inadequate revenues to support investments in warehouses and delivery vehicles to serve all local areas. This was thought to have contributed to poor delivery rates. In response, three alternative arrangements for supply logistics were introduced in 2012. Under these arrangements, the number of bid winners suppliers can select for deliveries within a county, was restricted.²² The first alternative delivery mode trialled was a medicine-tied distribution in which one distributor was selected for one medicine. The second was a recipient-tied distribution under which primary healthcare facilities selected one out of four distributors nominated by suppliers and chosen by government. The third was a recipient-medicine distribution under which primary healthcare facilities were served by one of two distributors selected by a supplier. The medicine-tied and the recipient-medicine distribution models seem designed to accommodate short-line logistics units, whilst the one-recipient-one-distributor model comes closer to the full-line logistics units this paper recommends. Yang et al.²² evaluation of these arrangements showed introduction of the one-distributor-one-medicine model led to a drop of 7.78 percentage points in delivery rates for rural facilities and even worse the one-distributor-one-recipient model led to a drop of 19.85 percentage points in delivery rates. These findings suggest that simply limiting the number of distributors (winners) did not solve the problem of unreliable supplies. Attempts to exploit economies-of-scale and scope from consolidation may inadvertently undermine delivery rates and lengthen lead times. This conclusion runs contrary to theory and historical evidence.

Limiting the number of logistics units effectively increases market size or business volumes (per unit), but it still requires distributors to make investments in logistics infrastructure for deliveries to rural and not just urban areas. Until that happens, delivery rates are likely to fall if the contracted distributors do not offer much in terms of geographical coverage. Concentration of contracted distributors in urban areas meant urban primary health facilities were not severely affected. However, delivery rates fell in rural areas with no alternative options. Rural facilities responded to erratic supplies by making larger requisition orders but then the negative impact of delivery delays is greater with larger orders.²² The key policy lesson here is: logistics units selected during competitive

tendering must be those capable of serving all intended health facilities. Although selective contracting with only 2 distributors capable of serving all primary healthcare facilities may appear as a near-monopoly situation, it may be enough to hedge against supply risks and support competition if the probability of disruptions is low (< 0.1) and healthcare payers are sensitive to both prices and outputs (lead times, service levels etc.). A more risk-averse health planner may find this argument unconvincing; in which case, efforts must be made, over time, to encourage (smaller) distributors with limited geographical coverage to grow or merge into the kind of logistics units capable of serving both rural and urban facilities. The alternative arrangement of allowing short-line subnational distributors to coexist with full-line national entities will erode the “profits beneath the zero” that logistics units need to cover costly investments for direct (last mile) delivery of all health commodities to all health facilities, rural or urban, within 72 h. That said, the administrative and transaction costs of switching between competing logistics units should be seen as costs to be incurred for providing choice and for competition to thrive. Besides, switching costs can be minimized by inviting and selecting logistics institutions who are similar in terms of their outputs or ability to serve all delivery points in the relevant market.

5. Conclusion

Reducing communicable and non-communicable disease burden in LMICs requires uninterrupted supplies of medicines and other health commodities. Assuring uninterrupted supplies of all health commodities, whenever and wherever they are needed, means these commodities must be available, acceptable, affordable and accessible. The logistics tasks of product selection, demand quantification, procurement and inventory management are therefore critical for the success of various vertical and horizontal health programs. Since warehousing and distribution costs constitute up to 16% of medicines budget in LMICs,¹ it is important to find ways of putting societal resources to its best possible use. One way of doing this is: reconfiguring or transforming current arrangements for inventory management in LMIC markets into one characterised by a limited number (2–10) of full-line national logistics institutions serving governmental, public or non-governmental actors. Such an arrangement is most suited for assuring aggregate commodity security in LMICs with relatively lower household and national incomes. Limited multiplicity should offer adequate insurance against disruptions related to storage, warehousing and distribution. However, other sources of supply risks (regulatory, political, financial [taxes, tariffs, currency conversion and interest rates etc.]) will have to be hedged using different strategies.^{23,24} There are limits to the risk-reduction benefits of multiplicity. And, for choice and competition, one doesn't need a large number of logistics units than is necessary or sufficient. A large number of competitors is no substitute for price- and output-sensitive purchasers.¹¹ Thus consolidation shouldn't progress to the point of monopoly and multiplicity doesn't mean having an infinite number of logistics institutions. If there exists a state of prudent multiplicity, it will be at the point where consolidation and multiplicity almost mean the same thing.

This paper set out to examine whether multiplicity in supply logistics (inventory management) is desirable – and if so, what form of multiplicity will help meet social objectives of assuring commodity security. The preceding sections described a form of multiplicity that offers the dual benefits of consolidation and multiplicity. Historical evidence from the US and Europe show consolidation towards limited multiplicity improved service delivery (i.e., national coverage for a full-line range of products) with little or no harm to competition. European and US private wholesalers, as logistics institutions, responded to price-sensitive demands and cost-containment incentives to adopt cost-reducing process innovations and, in the long run, exploit economies of scale and scope. The paper, however, leaves health planners in LMICs to make judgements as to whether the costs are worth the benefits of having 2–10 full-line national logistics institutions serving all sectors. If health planners favour multiplicity but not the

suggestion to build 2–10 competing units, they will have to make judgements as to whether additional multiplicity offers more benefits than it adds to costs. It is unlikely that health planners in LMICs have stacks of hard data to make such difficult choices, which means they should give themselves ample time and work with a longer-time planning horizon in their attempts to implement reform change. Having settled on N number of competing logistics units, the next challenge is transforming current structures within LMICs into the “state of prudent multiplicity” chosen.

The initial drive to do so may come from (a coalition of) entrepreneurs, who understand the essence of aggregate commodity security to population health, getting together to form a national full-line wholesalers out of smaller regional short-line wholesalers. The initial drive may also come from public-sector outsourcing or contracting out distribution of health commodities in urban (rural) areas to private wholesalers or non-governmental entities whilst channelling public efforts towards supply security in rural (urban) areas. An example is initiatives by VillageReach (a nongovernmental agency) in Mozambique. Another example is Uganda's National Medical Stores opting for in-house supply to district stores whilst outsourcing last mile delivery from district stores to third-party logistics providers who use trucks, bicycles, boats and even head-loading to supply entire districts within 15 working days.^{25,26} The drive for change may also come from the creation of a private (quasi-government) logistics institutions (similar to Zambia's Medical Stores Limited) and/or healthcare purchasers demanding lower prices for inventory management and higher output levels (for e.g., door-to-door last mile distribution within the shortest possible time).²⁵ Still, other government interventions may be necessary. In China, for instance, government introduction and nationwide implementation of Good Supply Practices (GSP) requirements in 2004 for private wholesalers led to a fall in the number of private wholesalers from 15,000 to 7445.¹ Whether this reduction in multiplicity was due to mergers and acquisitions, or some private wholesalers simply exiting the market is unclear. Perhaps exiting wholesalers in China did not have the capacity (capital) to meet these GSP requirements or were unwilling to incur the continuing or repeated outlays needed to meet these requirements, given expected business volumes. Both ways, the evidence indicates a trend towards consolidation. LMICs could therefore adopt a similar strategy with the explicit objective of reaping the benefits of both consolidation and multiplicity.

To sum up, this paper believes a balance between multiplicity and consolidation in logistics institutions or systems (at least with respect to the task of inventory management) can be achieved in LMICs. It will require health planners to find ways of turning existing logistics institutions operating in the public, private and even non-governmental sectors into types that are capable of assuring aggregate commodity security. Our ongoing research considers this issue, and we encourage theoretical and/or empirical work by others on what is arguably a less researched topic. These lines of inquiry could, for e.g., look at evaluating the performance of transformed logistics arrangements in LMICs in terms of how effective competition is, whether there is adequate hedging of disruption risks to support uninterrupted supplies, and the impact on health outcomes.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The statements, views and conclusions expressed in this paper are those of the author and not necessarily those of the author's affiliated institutions (in the past or present). The author takes full responsibility for all errors. No external research funding was used in support of this paper.

References

1. WHO. World Medicines Situation 2011: Storage and Supply Chain Management. Available at: <http://apps.who.int/medicinedocs/> 2018. Accessed May 16.
2. Raja R. *Commodity Security for Essential Medicines: Challenges and Opportunities*. Arlington, VA: USAID | DELIVER PROJECT. 2008. Available at: https://www.rhsupplies.org/uploads/tx_rhscpublications/CSEssMedChall.pdf. Accessed November 12, 2019.
3. Bornbusch A, Bates J. Multiplicity in public health supply systems: a learning agenda. *Glob Health Sci Pract* 2013;1:154.
4. McAfee RP, McMillan J. Competition and game theory. *JMR* 1996;33:263–267.
5. Baumol WJ. *Regulation Misled by Misread Theory: Perfect Competition and Competition-Imposed Price Discrimination*. Washington DC: The AEI Press. 2006.
6. Vickers J. Market power in competition cases. *Eur Compet J* 2006;2:3–14.
7. Clarke R, Davies SW. Market structure and price-cost margins. *Economica* 1982;49:277–287.
8. Tetteh EK. Implementing differential pricing for essential medicines via country-specific bilateral negotiated discounts. *Appl Health Econ Health Pol* 2009;7:71–89.
9. van Valen M, Jamieson D, Parvin L, Ramirez CL. Dispelling myths about drug procurement policy. *Lancet Glob Health* 2018;6:e609–e610.
10. Prosser W, Folorunso O, McCord J, et al. Redesigning immunization supply chains: results from three country analyses. *Vaccine* 2021;39:2246–2254.
11. Tetteh E. Creating reliable pharmaceutical distribution networks and supply chains in African countries: implications for access to medicines. *Res Soc Adm Pharm* 2009;5:286–297.
12. Lee HL. Aligning supply chain strategies with product uncertainties. *Calif Manag Rev* 2002;44:105–119.
13. Lee HL. The triple-a supply chain. *Harv Bus Rev* 2004;82:102–112.
14. Wieland A. Selecting the right supply chain based on risks. *J Manuf Technol Manag* 2013;24:652–668.
15. Fein AJ. Understanding evolutionary processes in non-manufacturing industries: empirical insights from the shakeout in pharmaceutical wholesaling. *J Evol Econ* 1998;8:231–270.
16. Clemons EK, Row M. McKesson drug company: a case study of Economost - a strategic information system. *J Manag Info Sys* 1988;5:36–50.
17. Oswald SL, Boulton WR. Obtaining industry control: the case of the pharmaceutical distribution industry. *Calif Manag Rev* 1995;38:138–162.
18. Kaiser Family Foundation. Follow the pill: Understanding the US commercial pharmaceutical supply chain. Available at: <https://www.kff.org/wp-content/uploads/2013/01/follow-the-pill-understanding-the-u-s-commercial-pharmaceutical-supply-chain-report.pdf> March 2005. Accessed March 12, 2020.
19. Walter E, Dragosits A, Said M. Access to pharmaceutical products in six European countries - analysis of different pharmaceutical distribution systems. *Farneconomia (Online)* 2012;13:33–41.
20. Andersson F. The distribution of pharmaceuticals in Europe - current and future trends in wholesaling. *Health Pol* 1994;27:271–292.
21. Taylor D, Mrazek M, Mossialos E. Regulating pharmaceutical distribution and retail pharmacy in Europe. In: Mossialos E, Mrazek M, Walley T, eds. *Regulating Pharmaceuticals in Europe: Striving for Efficiency, Equity and Quality*. Maidenhead: Open University Press; 2004. p. 196–212.
22. Yang L, Huang C, Liu C. Distribution of essential medicines to primary care institutions in Hubei of China: effects of centralized procurement arrangements. *BMC Health Serv Res* 2017;17:727.
23. Jaberidoost M, Nikfar S, Abdollahi A, Dinarvand R. Pharmaceutical supply chain risks: a systematic review. *DARU* 2013;21:69.
24. Chopra S, Sodhi MS. Managing risk to avoid supply-chain breakdown. *MIT Sloan Manag Rev* 2004;46(1):53–61. Fall Issue:53–61.
25. GAVI. Outsourcing the Distribution Component of Vaccine and Medicine Supply chains. Available at: <https://www.transaid.org/wp-content/uploads/2016/02/GAVI-Outsourcing-Report.pdf> 2021. Accessed September 25.
26. McCabe A, Seiter A, Diack A, Herbst CH, Dutta S, Saleh K. Private Sector Pharmaceutical Supply and Distribution Channels in Africa: A Focus on Ghana, Malawi and Mali. Available at: <https://openknowledge.worldbank.org/bitstream/handle/10986/13590/656010WP00PUBL00PvtSectorPharma0811.pdf?sequence=1&isAllowed=y> August 2011.
27. USAID. *The Logistics Handbook: A Practical Guide for the Supply Chain Management of Health Commodities*. 2nd ed. Arlington, VA: USAID | DELIVER PROJECT. 2011.(First edition 1998). Available at: <http://ghsupplychain.org/sites/default/files/2019-07/Logistics%20Handbook.pdf>. Accessed June 18, 2018.
28. USAID. Logistics System Assessment Tool (LSAT). Arlington, VA: USAID | DELIVER PROJECT. Available at: https://publications.jsi.com/JSIInternet/Inc/Common/_download_pub.cfm?id=14130&lid=3; Accessed May 8, 2018.